DOCUMENT RESUME

ED 298 771 FL 017 567

AUTHOR Gradman, Harry L.; Hanania, Edith
TITLE The Relation of Task to Performance in Testing Verbs.
PUB DATE Mar 88
NOTE 21p.; Revised version of a paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages (22nd, Chicago, IL, March 8-13, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cloze Procedure; College Students; Comparative Analysis; *English (Second Language); *Error Analysis (Language); Error Patterns; Higher Education; *Language Tests; Multiple Choice Tests; *Test Format; Test Reliability; *Verbs
IDENTIFIERS Indiana University Bloomington; Nonnative Speakers

ABSTRACT

A study investigated the variability of language performance on different types of testing task, global versus discrete-focus. Three tests (cloze, multiple-choice, and fill-in-the-blank) were developed to measure learners' knowledge of five verb forms. The tests, containing corresponding items designed to elicit equivalent structures, were administered to nonnative speakers of English grouped by proficiency level and by language background and also administered to a smaller group of native speakers. The results showed a clear pattern of variability, with students performing best on the multiple-choice task and least well on the cloze task, with greater variability at lower proficiency levels and on the more difficult verb structures. Differences in performance also seemed to be closely related to the production versus recognition features of the elicitation task. Analysis of two other factors, language background and error type, suggest a role for first language in performance variability. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED298771

Harry L. Gradman and Edith Hanania
Indiana University, Bloomington, Indiana, U.S.A.
Paper presented at the 22nd Annual TESOL Convention, 8-13 March 1988, Chicago, Illinois.

THE RELATION OF TASK TO PERFORMANCE IN TESTING VERBS

The work we are presenting here is an on-going investigation of language performance on different testing tasks, involving selected verb structures.[1]

It has been suggested that a learner's performance in the target language varies depending on whether the learner's attention is focused on meaning, as in a natural conversation, or on form, as in a grammar-based task. References to this subject in the literature have appeared in two main areas: in writings on the monitor model, and in interlanguage studies.

In the monitor model, Krashen (1981) considers that the learner may bring to bear on language production a knowledge of consciously learned rules, but only under certain conditions, these being: enough time; focus on form; and knowledge of the rules. In other words, when individuals focus on form, they monitor their language production by applying formally learned, consciously available rules. This notion has been used to interpret differences in the reported order of acquisition of morphemes, suggesting that data elicited through discrete-point tasks would yield a different order of acquisition than data obtained otherwise (Dulay, Burt and Krashen 1982).

---

[1] An account of our early work in this area appears in Individualizing the Assessment of Language Abilities, Eds. de Jong & Stevenson, Multilingual Matters, (in press).

FL017567

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. Gradman
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

Variability in performance has also become a focus of interest in interlanguage studies. Early work by Dickerson (1975) on acquisition of English by Japanese learners, and subsequent work by Tarone (1979 and 1982) substantiate the notion of variability along a continuum of styles, ranging from formal to communicative, the latter, according to Tarone, being the most systematic.

It is clear that such variability would have particular relevance to language testing. If discrete test items that focus on linguistic form invoke conscious knowledge of rules that may not have become part of the productive system of the learner, then global tests may reflect more accurately the learner's ability to apply those rules in communicative situations.

In the present stage of the work, we investigated the performance of a group of students on different types of testing tasks and examined the way in which performance varied with the testing task, the level of proficiency of the learners, and their language background. The area selected for testing was English verb forms. Verbs are a central part of English sentence structure, and the various verb forms are acquired at different stages of language learning. The work could therefore be expected to provide a rich body of data for comparative analysis.

PROCEDURE

Three tests were prepared to measure learners' knowledge of selected verb forms: cloze, multiple-choice, and fill-in-the-blank.

The three tests each consisted of thirty items which were designed to elicit corresponding verb structures. However, each of the tests represented a different type of task. Cloze required production of an appropriate verb within the context of continuous discourse, the attention of the test takers presumably being focused more on content than on form. With multiple-choice, the task was essentially one of recognition, requiring selection of the correct form of the verb from four alternatives. In fill-in-the-blank, the task involved production, as in cloze, but since the base form of the lexical verb was given, the focus of the production task was on form. In that respect, the fill-in-the-blank test was intermediate between the other two tests in the type of task it involved.

The subjects for this study were 213 nonnative speakers of English studying at Indiana University. They were students in the Intensive English Program and in three graduate linguistics classes, and they represented several language backgrounds which could be grouped under three main headings: Arabic, Asian, and Romance. The subjects were divided into three groups by proficiency level -- low, intermediate, and advanced -- on the basis of TOEFL scores. The low group consisted of students with scores below 420, the intermediate

group had students with scores between 420 and 530, and the advanced group had scores above 530.

The three tests were given at the same session in the order in which items on one test were considered to be least likely to affect the others: cloze, fill-in-the-blank, multiple choice. Enough time was allotted for all students to complete the test, and the papers were collected separately for each task. The tests were also given to 30 native speakers of English as a reference group.

The analysis of data was based on percent scores. The cloze test was scored for correct verb form. This was done regardless of lexical choice in the case of the simple past and simple present. Non-verb entries were considered inapplicable and were eliminated from the calculation. This method of scoring was used to insure that the cloze scores reflected only correct use of verb form, which was the concern of this investigation, and in that respect to make the cloze scores comparable with the scores from the other two tests.

The analysis focused on five specific verb structures: V-ed (simple past tense, both regular and irregular), V-s (present tense, 3rd person singular), BE (present,is/are), perfect, and modal, comprising twenty-one items in all. Mean scores on the three tests were computed for the structures combined and for individual structures. The tabulated data enabled comparison of performance across task for the whole group, as well as by level, by verb structure, and by language background.

RESULTS

Table 1 gives the mean percent scores for all verb structures (21 items), distributed by test type and by proficiency level. The figures show that the results for cloze and fill-in-the-blank tests are quite similar at each of the three levels, but that multiple-choice scores are significantly higher in all cases.

It is also evident that the extent of the differences varies according to level. Thus, score differences between multiple-choice and the other two tests are most pronounced for the weakest group of students and, as might be expected, differences are smallest for the advanced group.

The question arises as to whether the specific verb forms, taken individually, reflect similar overall differences. Table 2 gives the figures for each of the five specific verb forms, but for all levels combined.

The results here do reveal a pattern of variability. For the V-ed structure, there are hardly any differences among the tests. The differences between multiple-choice and the other two are more pronounced for the V-s and BE structures, and are even greater for the perfect and modal structures. It was particularly interesting for us to find that this sequence of increasing variability as one goes down the list of verb structures parallels the sequence of decreasing overall scores (that is, increasing difficulty) as can be seen in the last column of Table 2. Here, the mean scores for all

the tests combined show that the whole group performed best on V-ed, followed by BE and then V-s, and performed worst on perfect and modals. This observation suggests that a relationship may exist between patterns of variability in test scores and the difficulty of a particular verb structure. We will return to order of difficulty again later.

The next question to consider is -- How is this variability by structure related to the learners' level of proficiency? Table 3 shows how the test scores are distributed by verb structure and by proficiency level. Only three sets of data are given, as examples.

If we look at the easiest structure, V-ed, at the top, we find no significant difference in test scores regardless of the task at any of the three levels.

On the other hand, if we look at the more difficult structures examined, perfect and modal (only modal is shown in Table 3), we find that the multiple-choice scores are significantly higher than the other two, at all levels. Again the differences are greatest for the weakest group of students.

The middle of Table 3 gives results for the structure V-s, as an example of the two structures that are intermediate in difficulty, V-s and BE. Here there is mixed variability. Significant differences between multiple-choice and cloze appear at the two lower levels of proficiency for V-s (shown in the Table), and at the two upper levels of proficiency for BE (not shown in the Table).

At this point, it is useful to examine the variability in student performance in relation to the contrasting features that each of the testing tasks represents. Two sets of features are involved: global vs. discrete, and production vs. recognition. The relationships are illustrated in Table 4.

Cloze and fill-in-the-blank are both production tasks, but they contrast in the global vs. discrete feature. The differences in mean scores between these tasks are insignificant.

Fill-in-the-blank and multiple-choice tasks are both discrete, focusing on form, but they contrast in the production vs. recognition feature. Here the differences in mean scores are marked.

Cloze and multiple-choice contrast in both sets of features: global vs. discrete, and production vs. recognition. The differences here are even more marked.

This pattern of relationships suggests that, in variability of performance on testing tasks, the role of production vs. recognition may be more prominent than that of global vs. discrete focus.

Language Background

One major factor in variability of performance could be the learner's language background. We have therefore extended the analysis of data to investigate this aspect. For this purpose, it was convenient to subdivide the sample of 213 students into three main language groups: Arabic, Asian

(including Chinese, Japanese, Malay, and Indonesian), and Romance (including French, Italian, and Spanish). Since details of the work have already been described above, only a brief summary of our results to date will be presented here.

Table 5 gives the overall mean scores on the three tests, for each of the language groups, at the advanced, intermediate, and low proficiency levels. Let us first look for evidence of the overall patterns we have observed so far. Then we will identify a few noteworthy differences by language background.

On the whole, when we compare performance across tasks (the columns) we do observe the same general pattern as before. Scores are higher on multiple-choice than on the other two tasks. Scores on cloze and fill-in-the-blanks are more similar, and differences between tasks are more marked at the lower levels of proficiency.

However, other aspects of variability also appear when the data are examined for specific verb structures. In order to avoid extensive tabulations, only one example, the verb BE, is given in Table 6. Although, as can be seen, the subdivided samples are quite small in some instances, a number of interesting generalizations seem to emerge:

1. Asian students at the low proficiency level appear to have a marked advantage in the multiple-choice test over the other students. They attain higher scores on all the

verbs (except V-ed on which all students achieved high scores), a fact which may simply reflect an emphasis in their English language learning abroad on preparation for objective, grammar-oriented tests.

2. The Romance language group at low and intermediate levels has a higher cloze performance on BE relative to the two other groups. In this case, the advantage may reflect greater familiarity with the verb structure BE through the first language.
3. The Arabic group of advanced-level students shows a consistently weaker performance than the other two groups, particularly on BE. One possibility may be that there is a negative influence here from the first language, Arabic, where the copula is not used in the present tense.

On the basis of this preliminary study of the language background factor, it would appear that the overall patterns of variability are on the whole similar, but that there are notable differences in performance on specific structures as well as some variation in the relative advantage of one test type over another. Error analysis work presently in progress is expected to provide more detailed information and, hopefully, a more substantive interpretation of these differences.

Order of Difficulty

We would now like to touch upon two further points of interest that have emerged from this study. The first

concerns the question of order of difficulty of verb structures. While we do not claim that the small number of test items on the specific structures examined can lead to a definitive generalization about order of acquisition, we thought that it would be interesting to compare the relative order of difficulty of the structures, as it is reflected in student performance on each of these three tasks.

Table 7 summarizes the results. A very clear -- unexpectedly clear -- pattern emerges. The order of difficulty is very consistent for the cloze and fill-in-the-blank tasks. In contrast, the order is very erratic for the multiple-choice task. This finding is in accord with the view, expressed by Tarone and others, that certain types of data in language acquisition research are more systematic than others. In this case, language production data appear to be more reliable than recognition data. The point merits further investigation.

Error Analysis

The second line of interest concerns the extension of this research into the detailed analysis of the errors made on each of the tasks. This aspect of the work is expected to provide insights into language development. It would also throw light on the more complex nature of the differences between the three testing tasks.

To start with, we have carried out an analysis of errors on the multiple-choice task. To illustrate the type of

information we obtained, we present a summary of preliminary results obtained for V-ed and V-s structures:

1. Failure to inflect the verb accounts for only a small proportion of the error.
2. Ignorance of the rules of use of other structures in the distrators accounts for many more incorrect choices.
3. Lack of attention to cues <u>across</u> clauses is another major source of error, particularly at the low proficiency level. It appears that, even in a situation where monitoring is supposed to take place, the early learner has difficulty in processing longer sentence units.

The above observations, together with corresponding results which are now emerging from the analysis of the other verb structures, are beginning to yield patterns of difficulty in language acquisition that second language learners encounter in their performance on language testing tasks. Here again, the work is in progress.

CONCLUSIONS

To sum up, the purpose of the present work was to determine how performance on certain verb forms varies according to the type of task, and how this variation is affected by level of proficiency, and language background. On the basis of the results we have presented, the following conclusions may be drawn.

1. A pattern of variation in student performance emerges, the major differences being between a cloze type and a multiple-choice test.
2. The extent of this variation depends on the learners' level of proficiency, the lower levels showing the greatest differences.
3. The extent of the variation also depends on the particular verb structure involved, the structures we found to be more difficult showing greater differences.
4. Language background is a factor which introduces additional specific effects that tend to be superimposed on the general pattern of variability.
5. Differences in task performance appear to be closely related to the feature of language production vs. recognition. However, the observed variability may also reflect more complex differences in these tasks.

Notes

1. The passages were taken from the following books:

   Hill, L.A. Intermediate Stories for Reproduction. London: Oxford University Press, 1965.

   Royds-Irmak, D.E. Beginning Scientific English, Book 1. London: Nelson, 1975

REFERENCES

Dickerson, Lonna J. The Learner's Interlanguage as a System of Variable Rules. TESOL Quarterly, 9:4 (December 1975), 401-407.

Dulay, H., Burt,M. and Krashen, S. Language Two. New York: Oxford University Press, 1982.

Krashen, Stephen D. Second Language Acquisition and Second Language Learning. Oxford: Pergamon Press, 1981.

Tarone, Elaine E. Interlanguage as Chameleon. Language Learning, 29:1 (June 1979), 181-191.

Tarone, Elaine E. Systematicity and Attention in Interlanguage. Language Learning, 32:1 (June 1982), 69-84.

Table 1

Results by Level

(All Verb Structures)

		Mean Score%		
Level	N	CL	MC	FB
Adv.	69	85	94	87
Int.	79	69	83	67
Low	65	42	58	40
TOTAL	213	66	79	66

Differences are significant at $p < .01$ at all levels for MC/CL and MC/FB.

CL = cloze
MC = multiple-choice
FB = fill-in-the-blank

Table 2

Results by Verb Structure

(All Levels)

	Mean Score %			
Structure	CL	MC	FB	All Tests
V-ed	86	84	84	84
V-s	58	74	64	65
BE	64	83	75	74
Perfect	53	75	51	60
Modal	38	75	39	51

N = 213

Differences are significant for MC/CL at $p < .01$ for all structures except V-ed.
Differences are significant for MC/FB at $p < .01$ only for perfectives and modals.

V-ed = simple past tense, both regular and irregular
V-s = present tense, 3rd person singular
BE = present, is/are

Table 3

Results by Level for Three Selected Structures

Structure: V-ed		Mean Score %		
Level	N	CL	MC	FB
Adv.	69	98	96	96
Int.	79	92	87	85
Low	65	66	67	67

Differences are not significant at $p < .01$

Structure: V-s		Mean Score %		
Level	N	CL	MC	FB
Adv.	69	87	92	89
Int.	79	59	80	69
Low	65	26	47	33

Differences are significant at $p < .01$ for CL/MC at the two lower levels.

Structure: Modal		Mean Score %		
Level	N	CL	MC	FB
Adv.	69	63	90	68
Int.	79	34	76	39
Low	65	15	57	10

Differences are significant at $p < .01$ for all levels for MC/CL and MC/FB.

Table 4

Contrasting Features of the Tasks

		Differences in mean scores
CL/FB	production/production global/discrete	-
FB/MC	discrete/discrete production/recognition	+
CL/MC	global/discrete production/recognition	++

Table 5

Results by Level and Language Group

(All Verb Structures)

			Mean Score %		
Level	L.Group	N	CL	MC	FB
Adv.	Arabic	9	77	92	82
	Asian	42	85	93	86
	Romance	10	88	96	92
Int.	Arabic	26	64	77	62
	Asian	32	70	87	73
	Romance	18	74	86	64
Low	Arabic	47	39	56	39
	Asian	12	56	71	49
	Romance	6	35	48	29

Table 6

Results by Level, Language Group and Verb Structure

Structure: BE

			Mean Score %		
Level	L.Group	N	CL	MC	FB
Adv.	Arabic	9	50	100	83
	Asian	42	88	98	92
	Romance	10	80	100	95
Int.	Arabic	26	54	83	73
	Asian	32	66	87	82
	Romance	18	89	92	92
Low	Arabic	47	35	55	51
	Asian	12	50	83	54
	Romance	6	75	67	58

Mean TOEFL score by level and language group:

Int.	Arabic	453 <	Asian	472 =	Romance	472
Low	Romance	349 <	Arabic	375 <	Asian	395

Table 7

Order of Difficulty of Verb Structures

(All Languages)

Adv. Level	CL	MC	FB	All Tests
V-ed	1	2	1	1
V-s	2	4	3	3
BE	3	1	2	2
Perf	4	3	4	4
Modal	5	5	5	5
Int. Level				
V-ed	1	1	1	1
V-s	3	4	3	3
BE	2	2	2	2
Perf	4	3	4	4
Modal	5	5	5	5
Low Level				
V-ed	1	1	1	1
V-s	3	5	3	3
BE	2	2	2	2
Perf	4	4	4	4
Modal	5	3	5	5